REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	Technical	

4. TITLE AND SUBTITLE

A New Method of Edge Correction for Estimating the Nearest Neighbor Distribution

5. FUNDING NUMBERS

DAAL03-92-G-0322

6. AUTHOR(S)

Ernesto M. Floresroux and Michael L. Stein

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Chicago
Chicago, Illinois 60637

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

ARO 30167.10-MA

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Analysis of data in the form of a set of points irregularly distributed within a region of space usually involves the study of some property of the distribution of inter-event distances. One such function is G, the distribution of the distance from an event to its nearest neighbor. In practice, point processes are commonly observed through a bounded window, thus making edge effects an important component in the estimation of G. Several estimators have been proposed, all dealing with the edge effect problem in different ways. This paper proposes a new alternative for estimating the nearest neighbor distribution and compares it to other estimators. The result is an estimator with relatively small mean squared error for a wide variety of stationary processes.

14. SUBJECT TERMS			15. NUMBER OF PAGES
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

A New Method of Edge Correction for Estimating the Nearest Neighbor Distribution *

by

Ernesto M. Floresroux
and
Michael L. Stein

TECHNICAL REPORT NO. 385

Department of Statistics
The University of Chicago
Chicago, Illinois 60637

January 1994
Revised October 1994

* This research was supported in part by National Science Foundation grant DMS 92-04504 and Army Research Office grant ARO-92G0322. The first author acknowledges support from the Instituto de Matematicas, UNAM, through the DGAPA.

Computations for this document were performed using computer facilities supported in part by the National Science Foundation under grants DMS 89-05292, DMS 87-03942, and DMS 86-01732 awarded to the Department of Statistics at The University of Chicago, and by The University of Chicago Block Fund.

A New Method of Edge Correction for Estimating the Nearest Neighbor Distribution

Ernesto M. Floresroux

Michael L. Stein

Abstract

Analysis of data in the form of a set of points irregularly distributed within a region of space usually involves the study of some property of the distribution of inter-event distances. One such function is G, the distribution of the distance from an event to its nearest neighbor. In practice, point processes are commonly observed through a bounded window, thus making edge effects an important component in the estimation of G. Several estimators have been proposed, all dealing with the edge effect problem in different ways. This paper proposes a new alternative for estimating the nearest neighbor distribution and compares it to other estimators. The result is an estimator with relatively small mean squared error for a wide variety of stationary processes.

1. Introduction

Spatial point processes have been widely used as models in many scientific and technological fields including ecology and biology (Diggle, 1983), astronomy (Neyman and Scott, 1958), and archeology (Donnelly, 1978). Exhaustive sampling of patterns is becoming more commonplace because images can be easily digitized, and data sets in such a form are more readily available. In order to describe, analyze, and make inference on such patterns, properties of the distribution of inter-event distances are usually studied. There is considerable work in estimating the reduced second moment measure (Ripley, 1988; Stein, 1993), the empty space function (Baddeley and Gill, 1993), and the nearest neighbor distribution (Diggle, 1983). This work proposes a new way of estimating the nearest neighbor distribution function, $G(t)$, which is defined as the probability that a ball of radius t, centered at an arbitrary event of the process, contains at least one other event: this is equivalent to the probability that the distance between an arbitrary event and its nearest neighbor is less than or equal to t. Heuristically, by an arbitrary event we mean that we

have a large but finite number of events in some finite region, and one of these events is selected by simple random sampling. A precise definition can be found in Daley and Vere-Jones (1988).

An example of when G is a useful description of a spatial point process is the locations of trees in a forest (Diggle, 1983). A simple model for their locations would be a homogeneous Poisson process with intensity λ, for which $G(t) = 1 - \exp(-\pi\lambda t^2)$ in two dimensions. However, due to the nature of the process of seed dispersal, there may be more clumping of trees than would occur under the Poisson model. Alternatively, there may be a competitive relationship between trees that would cause trees to be more evenly spaced than under the Poisson model. The nearest neighbor distribution provides one method of distinguishing between these various possibilities and is particularly relevant when dependencies of the process over short distances are of interest.

Since mapped patterns are usually observed through windows with boundaries, events close to the boundary of the observation region might have their true nearest neighbor outside it. This makes edge effects an important component in the estimation of G, and any reasonable estimator should account for them. Several estimators have been proposed in the literature (Baddeley and Gill, 1993; Doguwa and Upton, 1990; Hanisch, 1984; Ripley, 1977), all of which deal with the edge effect problem in a different way. For observations near the boundary of study, the ball of radius t around these events is not entirely observed, rendering incomplete information about the nearest neighbor. Instead of discarding this information, a simple method for imputing the probability of an event in this ball conditioning on the data is proposed. It is based on considering other events in the observed part of the process for which a translation (if the process is stationary) and a rotation (if the process is stationary and isotropic) make it comparable in an appropriate sense to the problem event. The result is an estimator that has a smaller mean squared error than the two most commonly used estimators and its bias is of roughly the same order.

Furthermore, as opposed to Doguwa and Upton's estimator, it performs well in certain large sample scenarios where edge effects are kept approximately constant as the observed region grows. For this purpose, suppose Φ is a stationary point process on $\mathbb{R}^2$. If the observed area is a rectangle, keeping one side constant while letting the other one increase, edge effects will be severe even though the total area and the number of observed points tend to infinity. A second scenario where edge effects do not diminish occurs if instead of sampling only one window, the process is observed through a whole set of similarly sized windows. Baddeley, et. al. (1993) give an example for a three-dimensional spatial point process in which the events are locations of a certain feature

in monkey skulls and multiple windows arise because the point pattern is mapped in a number of well separated sampling volumes from the skull.

To study the performance of the estimators in different scenarios, three qualitatively different point processes covering a wide spectrum of alternatives are used as examples. As an example of an aggregated process, a Neyman-Scott cluster process (Neyman & Scott, 1958) is analyzed. For a process exhibiting some regularity or repulsion, a perturbed grid is chosen. In such a model, points on a grid are randomly moved according to a certain distribution. Finally, since no analysis is complete without studying the behavior under what is considered the canonical model for point processes, a homogeneous Poisson process is considered.

Section 2 defines some necessary notation. Section 3 reviews previously suggested estimators for the nearest neighbor distribution. Section 4 defines the new estimator. For simplicity, it is only defined for processes in $\mathbb{R}^2$, although the basic idea behind it applies in any number of dimensions. Section 5 provides a heuristic argument as to why our estimator should be preferred to the one suggested by Doguwa and Upton (1990). Section 6 describes the results of the simulation study.

2. Notation

Suppose Φ is a simple stationary point process in $\mathbb{R}^d$ and only a part of it is observed through a bounded window $W \subset \mathbb{R}^d$. As is customary in the literature, Φ will denote both a random set in $\mathbb{R}^d$ and a random measure. For any Borel set $A \subset \mathbb{R}^d$, the random variable $\Phi(A)$ counts the number of points of Φ that fall in A. Let $v_d(A)$ represent the $d-$dimensional volume of the set A; the subscript d will be suppressed if its meaning is unambiguous from the context. For any two points $P_1, P_2 \in \mathbb{R}^d$, the Euclidean distance between them is denoted by $d(P_1, P_2)$. Finally, $b(x,t)$ represents the ball of radius t around the point x.

The objective is to estimate G, the nearest neighbor distribution function for the underlying point process Φ. Denote the events in W by $P_1, \cdots, P_n$, so that $\Phi(W) = n$. For every event P_i, b_i denotes its distance to the nearest border of W and w_i the distance to its nearest neighbor inside W. That is, $w_i = \min_{j \neq i}\{d(P_i, P_j)\}$.

For all observed events P_i, Doguwa and Upton (1990) analyze the six possible orderings of the distance t at which we are estimating G, the distance b_i to the nearest border, and the distance w_i to the nearest neighbor. For five cases, it is known with all certainty whether the nearest neighbor is within t from the point of interest. This is not true in the case where $b_i < t < w_i$

(hereafter, this situation will be referred to as the *maybe* case), since the part of the circle that is not observed may or may not contain an event of the process. For all cases, including this last one, for large enough t, the balls of radius t around most events intersect the border, and thus are not completely observed.

Following Doguwa and Upton's (1990) notation, define the function $f(u, v)$ for $u, v \in \mathbb{R}$ as $f(u, v) = I\{u \leq v\}$, where I is an indicator function. For the observed window W, $W_{\ominus t}$ denotes the window W eroded by a ball of radius t, which can be written as $W_{\ominus t} = \{x \in W | b(x, t) \subset W\}$. Thus, $W_{\ominus t}$ is the set of points in W that are more than a distance t away from the border.

3. Existing Estimators

3.1 Reduced Sample Estimator

The most widely used estimator, usually referred to as the *reduced sample estimator,* was proposed by Diggle (1979) following Ripley (1977). For those events that are more than a distance t away from the boundary (all events P_i such that $t \leq b_i$; that is, $P_i \in W_{\ominus t}$), the whole circle of radius t around them is observed. Thus, with all certainty it can be said whether the nearest neighbor is within t. By selecting those events $P_i \in W$ for which this condition is met, that is, by allowing a guard area around the border, the sample gets restricted to events that satisfy only three cases out of the six possible orderings of the ranks of b_i, w_i, and t. As t grows, the size of the available sample decreases, and for large enough t, it will be zero. Formally, this estimator can be written as

$$\widehat{G}_1(t) = \frac{\sum_{i=1}^{n} f(t, b_i) f(w_i, t)}{\sum_{i=1}^{n} f(t, b_i)}.$$

This estimator of G counts the number of events in the eroded window $W_{\ominus t}$ for which the nearest neighbor is observed within t, and divides it by the total number of events in $W_{\ominus t}$. Although this method is intuitively appealing, it has certain drawbacks. Its range of estimation is quite limited: for example, in the case of a rectangle of sides s_x and s_y, the largest value of t for which the estimation is possible is $t \leq \min\{s_x, s_y\}/2$. This estimator sometimes renders a nonmonotone empirical distribution function, and for large t, the reduced sample estimator might be undefined.

3.2 Hanisch's Estimator

Hanisch (1984) proposed an alternative estimator for G. Instead of restricting the study to those events that lie in $W_{\ominus t}$, attention is focused on all objects for which the distance to the nearest neighbor is known. This is the subset of events $P_i \in \Phi \cap W$ for which $w_i \leq b_i$. This estimator

restricts itself to a different subset of events than the reduced sample estimator. It can be written as

$$\widehat{G}_2(t) = \frac{\sum_{i=1}^{n} f(w_i, t) f(w_i, b_i) p(w_i)}{\sum_{i=1}^{n} f(w_i, b_i) p(w_i)}.$$

A weight function p is added to account for the area within W in which any pair of events separated by a distance t could lie. In general, $p(z) = 1/v(W_{\ominus z})$, which renders $p(z) = (s_x - 2z)^{-1}(s_y - 2z)^{-1}$ for a rectangle of sides s_x and s_y. The purpose of this function is to make $\widehat{G}_2$ a ratio-unbiased estimator of G: that is, the ratio of expectations of the numerator and denominator is equal to the true nearest neighbor distribution. Such a property is often encountered in estimators of this kind.

Stoyan, Kendall & Mecke (1987, p. 128) proposed another version of this estimator that does not take into account the weight function. We made a comparative analysis of these two versions and we found no reason for not using p, except for a minor increase in computational effort, and ratio-unbiasedness is lost by doing so.

The estimator $\widehat{G}_2$ is well defined if at least one event in the study region has its nearest neighbor closer to it than the border. Conditional on the observed process, the denominator is constant, and thus the estimator is nondecreasing.

3.3 Doguwa and Upton's Estimator

Doguwa and Upton (1990) proposed another estimator that avoids some of the problems encountered by $\widehat{G}_1$ and $\widehat{G}_2$. It is well defined unless $\Phi(W) = 0$, and it is monotone in t. Most importantly, instead of just considering those events P_i for which only certain orderings of the distances t, b_i, and w_i are satisfied, it accounts for all the events simultaneously. For an event P_i for which $b_i < t < w_i$ (the *maybe* case), an ad hoc method of imputation is proposed. Figure 1 illustrates the problem solved by this method. It works well in certain cases, but it makes an inappropriate correction when the process is badly non-Poisson.

To construct such an estimator, the number of events $P_i \in \Phi \cap W$ for which $w_i \leq t$ is counted. This is known with all certainty in five out of the six ordering of t, b_i, and w_i. For the *maybe* case, only part of the ball of radius t around such an event is observed in W, and no other events of the process lie in it. Denote by $I_{P_i}(t) = b(P_i, t) \cap W$ the part of $b(P_i, t)$ that is inside W, and by $O_{P_i}(t) = b(P_i, t) \cap W^c$ the part outside W. The objective is to estimate $H(t; P_i) = \mathrm{P}\left\{\Phi\big(O_{P_i}(t)\big) > 0 | P_i, \Phi\big(I_{P_i}(t)\big) = 1\right\}$, the conditional probability of finding at least one event in the unobserved part of the ball of radius t around the event P_i given that there are no other events in the observed part.

If the process is assumed to be homogeneous Poisson, this imputation is trivial since the numbers of events of the process in disjoint regions are independent. Let $v_0 = v(W)$, and let $v_i = v\{I_{P_i}(t)\}$ and $h_i = v\{O_{P_i}(t)\}$ be the areas of the observed and unobserved parts of $b(P_i, t)$ respectively; clearly, $v_i + h_i = \pi t^2$ in two dimensions. For a Poisson process with intensity λ, $H(t; P_i) = 1 - \exp(-\lambda h_i)$. Estimating λ by the standard estimator $\hat{\lambda} = n/v_0$ yields $\widehat{H}(t; P_i) = 1 - \exp\{-\hat{\lambda} h_i\}$.

This probability has to be estimated for every event P_i for which $b_i < t < w_i$. Notice that t is considered fixed, and the set of events for which $H(t; P_i)$ has to be estimated is different for different values of t. For both small and large values of t, the number of events in the *maybe* case is zero.

¿From this, Doguwa and Upton (1990) recommend

$$\widehat{G}_3(t) = \frac{1}{n}\sum_{i=1}^{n} f(w_i, t) + \frac{1}{n}\sum_{i=1}^{n}\{1 - f(w_i, t)\}\{1 - f(t, b_i)\}\widehat{H}(t; P_i).$$

This result is independent of whatever is observed in $I_{P_i}(t)$, which makes estimating G by $\widehat{G}_3$ easy because only areas of parts of circles have to be calculated. However, in certain cases it can introduce serious bias problems if the process is sufficiently non-Poisson.

3.4 Other Estimators

Another estimator for G recently proposed by Baddeley and Gill (1993) treats edge effects as a censoring problem. Based on the analogy with censored survival data, the distance from an event to its nearest neighbor is taken to be right-censored by its distance to the border. By extending the Kaplan-Meier estimator and computing a cumulative hazard function, they construct a ratio-unbiased estimator that has the following expression:

$$\widehat{G}(t) = 1 - \prod_{r}\left(1 - \frac{\sum_{i=1}^{n} I\{w_i = r\} f(w_i, b_i)}{\sum_{i=1}^{n} f(r, w_i) f(r, b_i)}\right),$$

where r in the product ranges over those elements of $\{w_1, \ldots, w_n\}$ that are at most t. Baddeley and Gill (1993) show that this estimator has advantages over the reduced sample estimator $\hat{G}_1$. A direct comparison with the estimator suggested here remains to be done.

The estimators $\widehat{G}_1$ and $\widehat{G}_2$ can be combined rendering another way of estimating G. This can be done by considering

$$\widehat{G}(t) = \frac{\sum_{i=1}^{n}\Big\{f(t,b_i)f(w_i,t) + f(w_i,t)f(w_i,b_i)p(w_i)\Big\}}{\sum_{i=1}^{n}\Big\{f(t,b_i) + f(w_i,b_i)p(w_i)\Big\}}.$$

This combination yields an estimator that performs better than $\widehat{G}_1$ and $\widehat{G}_2$ both in terms of bias and mean squared error. This is due mainly to the fact that each estimator uses different sets of the orderings of t, b_i, and w_i. Though trivial, this seems to have been overlooked in the literature.

4. A New Estimator

Based on Doguwa and Upton's idea, this section presents another estimator for G. The imputation problem is approached with fewer and less restrictive assumptions. As before, the goal is to estimate $H(t;P_i)$, the probability of finding at least one event in the unobserved part of $b(P_i,t)$ if $b(P_i,t) \not\subset W$. For simplicity, we describe the estimator for processes on $\mathbb{R}^2$, although the same approach can be used in any number of dimensions.

For this purpose, the *translation* A_u of a set $A \in \mathbb{R}^2$ for $u \in \mathbb{R}^2$ will be defined as $A_u = A+u = \{y+u | y \in A\}$. The *rotation* of a set $A \in \mathbb{R}^2$ by an angle θ is defined as $R_\theta A = \{R_\theta y | y \in A\}$, where R_θ is the orthonormal matrix that rotates points in $\mathbb{R}^2$ counterclockwise by an angle θ.

For Φ a stationary simple point process, $t > 0$ fixed, $P_i \in \Phi \cap W$, and $P_j \in \Phi \cap W_{\ominus t}$, P_j will be said to be *analogous* to P_i if $\Phi\{b(P_j,t) \cap W \cap W_{P_i-P_j}\} = 1$. In other words, P_j is analogous to P_i if the only event of the translated process that is in the part of $b(P_j,t)$ intersecting $W_{P_i-P_j}$ is P_j itself. This is illustrated in Figure 2(c).

If Φ is also assumed to be isotropic, then the definition of *analogous* points can be extended. For $\theta \in (0,2\pi]$, an event P_j is *analogous at* θ to P_i if $\Phi\{b(P_j,t) \cap W \cap R_\theta W_{P_i-P_j}\} = 1$. Then, *analogous* as defined for stationary processes can be stated as being *analogous at 0.* This is illustrated in Figure 2(d).

For $P_i, P_j \in \Phi \cap W$, where Φ is a stationary simple point process, define

$$T_i(P_j) = \begin{cases} 1 \text{ if } P_j \text{ is analogous to } P_i \\ 0 \text{ otherwise.} \end{cases}$$

Lastly, for $P_i \in \Phi \cap W$ and $P_j \in \Phi \cap W_{\ominus t}$, where Φ is a stationary and isotropic simple point process, define $\theta_i(P_j)$ as the *total angle of rotation* that makes P_j analogous to P_i. This can

be written as $\theta_i(P_j) = \int_0^{2\pi} I\{\text{ analogous at } \theta\}d\theta$. This implies that if $\Phi\big(b(P_j,t)\big) = 1$. that is, the nearest neighbor of P_j is not within t. then $\theta_i(P_j) = 2\pi$. The angles of rotation at which P_j is analogous to P_i do not necessarily form a simple arc.

To construct the estimator. we will first assume that the underlying process is simple and stationary. As before. $H(t; P_i)$ denotes the probability of finding at least one event in the unobserved part of $b(P_i, t)$. Instead of acting as if the process were Poisson, this probability can be estimated by

$$\widehat{H}_*(t; P_i) = \frac{\sum_{j=1}^n T_i(P_j) f(w_j, t)}{\sum_{j=1}^n T_i(P_j)}.$$

This estimator looks for the events in the observed realization of the process Φ that are analogous to P_i, assesses for such events whether an event of the process lies in the part of the circle corresponding to the unobserved part of $b(P_i, t)$, and computes the ratio of these two counts. Although $\widehat{H}(t; P_i)$ might be undefined if there are no analogous points to P_i, if at least one such point is observed, this conditional probability can be estimated. If there are no analogous points we will take $\widehat{H}_*(t; P_i) = 0$. The new estimator then takes the same form as Doguwa and Upton's replacing $\widehat{H}(t; P_i)$ with $\widehat{H}_*(t; P_i)$.

To estimate $H(t; P_i)$ this way the only assumption required was stationarity of the process. If it is believed that the process is also isotropic, a given realization of Φ might provide even more information. An event P_j might not be analogous to P_i in the first sense defined above. but there might exist a rotation of the process that makes it analogous to P_i. To see that this can be the case. consider P_2 in Figure 2(a). Figure 2(c) illustrates why P_2 is not analogous to P_1. However, if the orientation is changed. it can be seen that there is at least one rotation of Φ that makes it analogous to P_1. This rotation is depicted in Figure 2(d).

Apparently, assuming isotropy increases the number of analogous situations dramatically. Nevertheless, analogous points must be counted in a different way than before. In this case, it seems reasonable to define

$$\widehat{H}_{**}(t; P_i) = \frac{\sum_{j=1}^n \theta_i(P_j) f(w_j, t)}{\sum_{j=1}^n \theta_i(P_j)}.$$

where $\theta_i(P_j)$ is the angle of rotation of I_{P_i} around P_j as defined above. This means that if P_j is analogous to P_i and its nearest neighbor is not within t of it, $\theta_i(P_j) = 2\pi$ and $\theta_i(P_j)f(w_j, t) = 0$, whereas if it is, the angle of rotation is added to both the numerator and the denominator of $\widehat{H}_{**}(t; P_i)$.

Since computing angles can become quite a cumbersome task, this situation can be approximated by doing the following. For every event P_j, the translated part of the circle $(I_{P_i}(t))$ can be rotated a finite number of times m. That is, for every point P_j, m different values for $T_i(P_j)$ are obtained. Denote them by $T_i^k(P_j)$, for $k = 1, \ldots, m$. The counting argument for calculating $\widehat{H}_{**}(t; P_i)$ still applies, and it can be approximated by

$$\widehat{H}_{**}(t; P_i) = \frac{\sum_{j=1}^{n} \sum_{k=1}^{m} T_i^k(P_j) f(w_j, t)}{\sum_{j=1}^{n} \sum_{k=1}^{m} T_i^k(P_j)}.$$

¿From simulation studies it was observed that m need not be large. The difference between $m = 1$ and $m = 4$, though nonnegligible, is small. Between $m = 4$ and $m = 8$, this difference is hardly noticeable. Choosing $m = 4$ is convenient for rectangular W.

5. Discussion

The basic problem with estimating $H(t; P_i)$ as suggested by Doguwa and Upton (1990), by acting as if the process were Poisson, is that it the resulting estimator of $H(t; P_i)$ can be substantially biased if the process is not Poisson. Our procedure is designed to avoid this problem. More specifically, by letting the observation window W grow in an appropriate way, $\widehat{H}_*(t; P_i)$ converges to $H(t; P_i)$, whereas $\widehat{H}(t; P_i)$ converges to $1 - \exp\Big(-\lambda\, v\big(b(P_i, t) \cap W^c\big)\Big)$, which does not in general equal $H(t; P_i)$. However, it is awkward to give a precise meaning to these statements since the definition of $H(t; P_i)$ depends on W. Consider the following related problem, which does have a clear interpretation. For a fixed set A containing 0, we want to estimate

$$\eta(A) = P\big(T^*(0, A) \mid T(0, A)\big)$$

where

$$T(x, A) = \big\{\Phi(\{x\}) = \Phi(A_x \cap b(x, t)) = 1\big\} \quad \text{and}$$
$$T^*(x, A) = T(x, A) \cap \big\{\Phi(b(x, t)) > 1\big\}.$$

Roughly speaking, we want to think of $\eta(A)$ as $H(t; P_i)$, with 0 as P_i, $b(0, t) \cap A$ as I_{P_i}, and $b(0, t) \cap A^c$ as O_{P_i}. Doguwa and Upton are essentially estimating $\eta(A)$ by $1 - \exp\Big(-\hat{\lambda}\, v\big(b(0, t) \cap A^c\big)\Big)$, while we are using the empirical estimator of the conditional probability given by

$$\sum_{P_i \in W_{\ominus t}} I\{T^*(P_i, A)\} \Big/ \sum_{P_i \in W_{\ominus t}} I\{T(P_i, A)\}.$$

Then, assuming the region W grows in an appropriate sense and Φ is ergodic (Daley and Vere-Jones, 1988, p. 335), our estimator converges almost surely to $\eta(A)$, whereas Doguwa and Upton's converges almost surely to the generally incorrect $1 - \exp\Big(-\lambda\, v(b(0,t) \cap A)\Big)$. This suggests that, unlike Doguwa and Upton's approach, the proposed imputation procedure is a sensible way of calculating the conditional probability $H(t; P_i)$, whether or not the underlying process is Poisson. The simulations in the next section show that the bias in Doguwa and Upton's approach to estimating $H(t; P_i)$ can lead to substantial bias in estimating G in situations where our approach does not have this problem.

6. Simulation Results

This section presents a comparative simulation study of the estimators $\widehat{G}_i$, $i = 1, \ldots, 4$ for the nearest neighbor distribution in order to evaluate their performance under different types of stationary and isotropic point processes. The estimator $\widehat{G}_4$ will be taken as the non-rotating version of the estimator. Simulations considering the rotating version were run, yielding only a slight improvement over the non-rotating one, so we report only the results obtained from the non-rotating version. All the code was written in C following ANSI standards, and the simulations were run on a Sparc 2-Sparc 10 computer system.

Simulations were run in order to study the behavior of the estimators in a specific large sample scenario. Though several large sample scenarios can be explored in the theory of point processes, since the estimators described above try to account for edge effects, it is of value to see how they perform if the sample size grows and edge effects remain roughly constant. That is, it is desirable to maintain the proportion of events close to the border constant and large. If the region W grows, for example, by letting both sides of a rectangle grow or by letting the radius of a circular region go to infinity, edge effects become negligible. In such cases, all the estimators described above perform well in terms of bias and variance, and the differences between them are small.

In order to maintain constant edge effects, for the simulations we will consider processes defined throughout $\mathbb{R}^2$ that are observed through several nonintersecting windows, and these windows are assumed to be sufficiently apart so that the observed realizations of the processes in each one of them are essentially independent. Simulations were run for one through five observation windows.

For every process, several thousand replications (N) were run, and the different estimators were calculated for each one. The value of the $i-$th estimator of G in the $j-$th realization is

denoted by $\widehat{G}_i^j$. By a replication we mean that the process was simulated in the relevant number of windows m ($m = 1, \ldots, 5$) and $G(t)$ and $H(t; P_i)$ were calculated taking $n = \sum_{i=1}^{m} \Phi(W_i)$, which is the total number of observed events in the m windows.

For every process considered, several summary statistics were calculated: the estimated bias, the average squared error, and the difference of the absolute errors *. Since the estimated bias is usually small, the average squared error curves are in most cases a good approximation to the variance of the estimators. The difference of the absolute errors makes small differences between the estimators more noticeable than when the average squared error is used. $\mathrm{DAE}_{i,j}(t)$ will be positive for those values of t for which $\widehat{G}_i$ is performing, on average, better than $\widehat{G}_j$, and negative otherwise.

6.1 Parent-Offspring Process

Poisson cluster processes, first suggested by Neyman and Scott (1958) as a possible way of describing cosmological data, incorporate an explicit form of spatial clustering. A specific Neyman-Scott process was considered, where the parent events form a Poisson process with intensity $\rho = 10$, and the distribution of offspring P_M is given by $M = (2, 3, 4)$ with probabilities $P_M = (0.5, 0.25, 0.25)$. The offspring are then placed around their parent following a bivariate normal distribution with covariance matrix $\sigma^2 I$, with $\sigma = 0.05$. For simulation purposes, it is worth pointing out that parents outside the observation region can produce offspring inside W, so the process has to be simulated on a region larger than W.

Figure 3 shows the estimated bias for $\widehat{G}_i$ $i = 1, \ldots, 4$ for N =25,000 replications of the cluster process simulated on unit square windows. The expected number of events per unit area is 27.5. The bias curves of all four estimators follow a somewhat similar pattern. $\widehat{G}_3$ has the largest bias, and in a small range of distances $\widehat{G}_4$ is the only one with a positive bias. The average squared error curves (not shown) are all of about the same order, though $\widehat{G}_4$ always shows a smaller average squared error than $\widehat{G}_1$ and $\widehat{G}_2$. Except for large distances, this is also true for $\widehat{G}_3$ when compared to $\widehat{G}_1$ and $\widehat{G}_2$. $\widehat{G}_1$ has the largest ASE. The average difference of the absolute error curves enlarge these differences, showing that $\widehat{G}_4$ performs uniformly better than $\widehat{G}_1$ and $\widehat{G}_2$. For small distances,

---

* For a fixed value of t, the computing formulas used were;

Estimated bias: $\mathrm{EB} = \frac{1}{N} \sum_{j=1}^{N} \{\widehat{G}_i^j(t) - G(t)\}$.

Average squared error: $\mathrm{ASE}_i(t) = \frac{1}{N} \sum_{j=1}^{N} \{\widehat{G}_i^j(t) - G(t)\}^2$.

Difference of the absolute errors: $\mathrm{DAE}_{i,j}(t) = \frac{1}{N} \sum_{k=1}^{N} \{|\widehat{G}_i^k(t) - G(t)| - |\widehat{G}_j^k(t) - G(t)|\}$.

$\widehat{G}_3$ performs somewhat better than $\widehat{G}_4$, and the trend gets reversed as the distance t increases. This can be appreciated in Figure 4, where the performance of the estimators in the multiple window scenario (2, 3, 4, and 5 unit square windows) is shown. The bias of $\widehat{G}_1$, $\widehat{G}_2$, and $\widehat{G}_4$ decreases in every case, whereas $\widehat{G}_3$ has a nontrivial bias that does not diminish, regardless of the number of windows. $\mathrm{DAE}_{1,4}$ and $\mathrm{DAE}_{2,4}$ decrease, due mainly to the increase in the sample size and that both $\widehat{G}_1$ and $\widehat{G}_2$ are ratio unbiased. On the other hand, $\mathrm{DAE}_{3,4}$ shows the opposite trend, since its negative part gets closer to zero, and the range of positive differences gets larger as the number of windows increases, which means the relative performance of $\widehat{G}_4$ improves with respect to that of $\widehat{G}_3$. For five windows, except at very short distances, $\mathrm{DAE}_{3,4}$ is positive, suggesting that as the number of windows increases, $\widehat{G}_4$ becomes a uniformly better estimator than $\widehat{G}_3$.

6.2 A Regular Process

Several models for processes that exhibit some regularity have been proposed in the literature. In order to investigate the behavior of the estimators in this case, a perturbed grid was chosen. First, a grid where the events are separated by a distance d in each direction is constructed. This grid is randomly placed on the plane after rotating it by an angle θ chosen uniformly in the interval $(0, 2\pi]$. Then every event is moved randomly from its position — hence perturbed — independently from the other events. The distance each event moves follows a bivariate normal distribution with covariance matrix $\sigma^2 I$ centered at their unperturbed position. The parameters of the grid considered for this section were $d = 0.15$ and $\sigma = 0.02$. The final pattern is definitely regular, but it is not apparent, given the expected number of events per observation window ($1/0.15^2 \approx 44$), that the original process was an evenly spaced grid. Given the time required to run this process on the multiple window scenario, for one and two windows, N =25,000 replications were run; for three, N =15,000, and for four and five, N = 7500.

Figure 5 shows the bias of $\widehat{G}_3$ and $\widehat{G}_4$ for the multiple window scenario. The shaded regions represent approximate 95% confidence intervals for the bias at distance t. $\widehat{G}_4$ appears to be unbiased, except at the the range $0.14 < t < 0.17$, where there exists a small positive bias that seems to decrease as the number of windows increases. On the other hand, $\widehat{G}_3$ has a nontrivial bias that does not change with the number of windows. The estimators $\widehat{G}_1$ and $\widehat{G}_2$ (bias curves not shown) are for all practical purposes unbiased. Figure 6 shows the average squared error for the four estimators. It can be appreciated that overall, $\widehat{G}_4$ has the smallest average squared error of the four estimators. The behavior of $\widehat{G}_3$ is due mainly to the estimation of a nontrivial positive

value for H for events in the *maybe* case: the true value of this probability in this case is very close to zero.

6.3 Poisson Processes

In the case of the Poisson process, it will not be surprising to see that for almost every t, $\widehat{G}_3$ performs better than the other three estimators. A Poisson process with intensity $\lambda = 20$ was simulated on the unit square. For each number of windows (one through five), N =25,000 replications were run. The patterns the estimated bias and average squared error curves show do not change qualitatively with the number of windows. Hence, Figure 7 shows these two curves for ony one window. $\widehat{G}_1$ consistently has a larger bias than $\widehat{G}_2$ and $\widehat{G}_3$. It is noticeable that $\widehat{G}_2$ has overall the smallest bias. $\widehat{G}_4$ has a small bias, but it is the only estimator that shows a positive bias in some range. $\widehat{G}_1$ has the largest average squared error, which becomes indistiguishable from that of $\widehat{G}_2$ as the number of windows increases. As expected, $\widehat{G}_3$ has overall the smallest average squared error: that of $\widehat{G}_4$ lies between ASE_3 and ASE_2.

Figure 8 shows the average of the difference of absolute error curves. For all t, $\mathrm{DAE}_{1,4}$ and $\mathrm{DAE}_{2,4}$ are positive. Hence, this criterion indicates that on average $\widehat{G}_4$ performs better than $\widehat{G}_1$ and $\widehat{G}_2$. On the other hand, $\mathrm{DAE}_{3,4}$ is negative for almost every t (except when t is large), indicating that $\widehat{G}_3$ does a better job than $\widehat{G}_4$ estimating the nearest neighbor distribution if the underlying process is Poisson.

7. Conclusion

The three types of processes studied in the previous sections were chosen in order to analyze and compare the behavior of four estimators of the nearest neighbor distribution under diverse alternatives. If the process is Poisson or very nearly Poisson, $\widehat{G}_3$ will be the best estimator because of the way the *maybe* case is handled: the unknown probability is imputed with the approximate corresponding Poisson probability, which in these cases will be very close to the truth. On the other hand, by imputing using a more empirical approach, $\widehat{G}_4$ does nearly as well in these circumstances. That is, $\widehat{G}_4$ does not depend on a Poisson assumption for its edge correction to make sense, and that is the reason that in non-Poisson processes, where $\widehat{G}_3$ sometimes performs badly, $\widehat{G}_4$ performs well. In terms of mean squared error, in all three processes considered $\widehat{G}_4$ performs uniformly better than $\widehat{G}_1$ and $\widehat{G}_2$, and the bias of the three is of the same order. Thus, overall, $\widehat{G}_4$ appears to be a better estimator than $\widehat{G}_1$, $\widehat{G}_2$, and $\widehat{G}_3$.

References

Baddeley, A.J. & Gill, R.D. (1993). Kaplan-Meier estimators of interpoint distance distributions for spatial point processes. Preprint Nr. 718, Department of Mathematics, University Utrecht.

Baddeley, A.J., Moyeed, R.A., Howard, C.V., Reid, S. & Boyde, E. (1993). Analysis of a three-dimensional point pattern with replication. *Appl. Statist.* **42** 641–668.

Bartlett, M.S. (1974). The statistical analysis of spatial pattern. *Adv. Appl. Probab.* **6** 336-358.

Daley, D.J. & Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes.* New York: Springer Verlag.

Diggle, P.J. (1979). On parameter estimation and goodness-of-fit testing for spatial point patterns. *Biometrics* **35** 87–101.

Diggle, P.J. (1983). *Statistical Analysis of Spatial Point Patterns.* London: Academic Press.

Doguwa, S.I. & Upton, G.J.G. (1990). On the estimation of the nearest neighbour distribution, $G(t)$, for point processes. *Bio. J.* **32** 863-876.

Donnelly, K. (1978). Simulations to determine the variance and edge effect of total nearest neighbour distance. In *Simulation Methods in Archaeology*, I.R. Hodder, ed. Cambridge University Press, 91-95.

Hanisch, K.-H. (1984). Some remarks on estimators of the distribution function of nearest neighbour distance is stationary spatial point processes. *Math. Oper. Statist. ser. Statist.* **15** 409-412.

Neyman, J. & Scott. E.L. (1958). Statistical approach to problems of cosmology. *J. Roy. Statist. Soc. B* **20** 1-43.

Ripley, B.D. (1977). Modelling spatial pattern. *J. Roy. Statist. Soc. B* **39** 172-212.

Ripley, B.D. (1988). *Statistical Inference for Spatial Processes.* Cambridge University Press.

Stein, M.L. (1993). Asymptotically optimal estimation for the reduced second moment measure of point processes. *Biometrika* **80** 443-449.

Stoyan, D., Kendall, W.S. & Mecke J. (1987) *Stochastic Geometry and Its Applications.* New York: Wiley.

Figure 1. P is an event for which $b(P,t)$ is not fully contained in W. $I_P(t)$ and $O_P(t)$ (shaded area) represent the parts of $b(P,t)$ that are inside and outside W respectively. Given that $I_P(t)$ contains only one event of the process (P itself), the objective is to estimate the probability of there being at least one event in $O_P(t)$.

Figure 2. Denote by H_a the part of $b(P_1,t)$ that lies outside the window W, and v_a the part of it that lies inside. Since in **(a)** v_a is empty, P_1 is in the *maybe* case as described in the text. In **(b)**, it can be seen that P_4 is analogous to P_1 because v_b only contains P_4, whereas in **(c)** it can be observed that P_2 is not because P_3 is also in v_c. In **(d)**, by rotating $b(P_2,t)$ by approximately 157° counterclockwise, P_2 becomes analogous to P_1. In fact, there exists a continuum of angles for which this is case.

Figure 3. This figure shows the estimated bias of the four estimators when the Neyman-Scott process described in the text is observed through multiple windows. For each number of windows, $N = 25,000$ replications were run. $\widehat{G}_1$, $\widehat{G}_2$, and $\widehat{G}_4$ are plotted using the same vertical scale. Since the bias of $\widehat{G}_3$ is larger than the other biases by one order of magnitude, it is plotted separately.

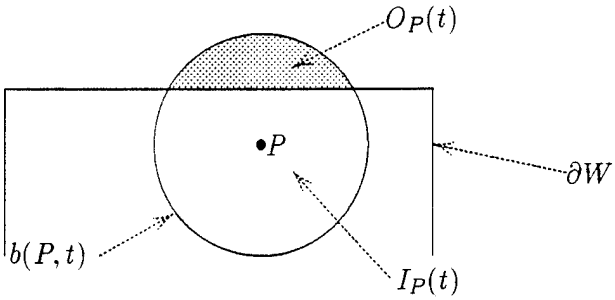
Figure 4. These plots show the average difference of the absolute errors for $\widehat{G}_1$, $\widehat{G}_2$, and $\widehat{G}_3$ when compared to $\widehat{G}_4$ when the Neyman-Scott process described in the text is observed through two, three, four, and five independent unit square windows.

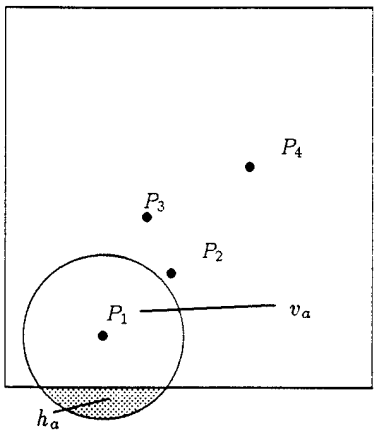
Figure 5. These figures show the average bias curves of $\widehat{G}_3$ and $\widehat{G}_4$ in the multiple window scenario for the perturbed grid. The shaded areas represent 95% confidence intervals for the average bias. Notice that the vertical scales are not the same.

Figure 6. These plots depict the average squared error curves of the four estimators when the perturbed grid is observed through multiple windows (one through five). All estimators have their ASE's in the same range.

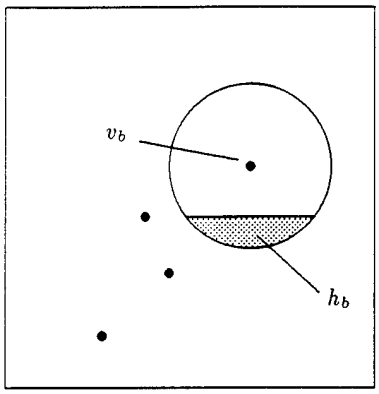
Figure 7. For one observation unit square window, the bias and average squared error of the four estimators when the process is Poisson are shown.

Figure 8. These plots show the average difference of the absolute errors obtained by comparing $\widehat{G}_1$, $\widehat{G}_2$, and $\widehat{G}_3$ to $\widehat{G}_4$ for simulations on one through five windows when the process is Poisson. All three plots have the same vertical scale.


$O_P(t)$
P
∂W
$b(P,t)$
$I_P(t)$

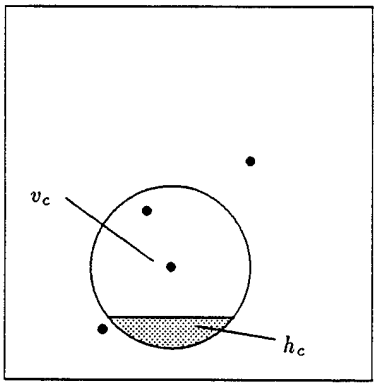


(a)



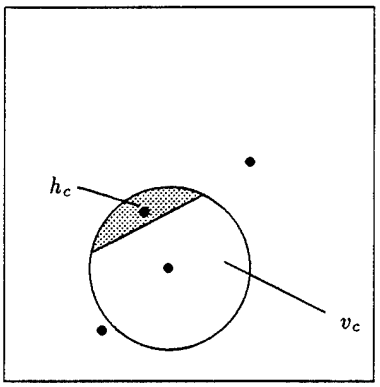
analogous

(b)



not analogous

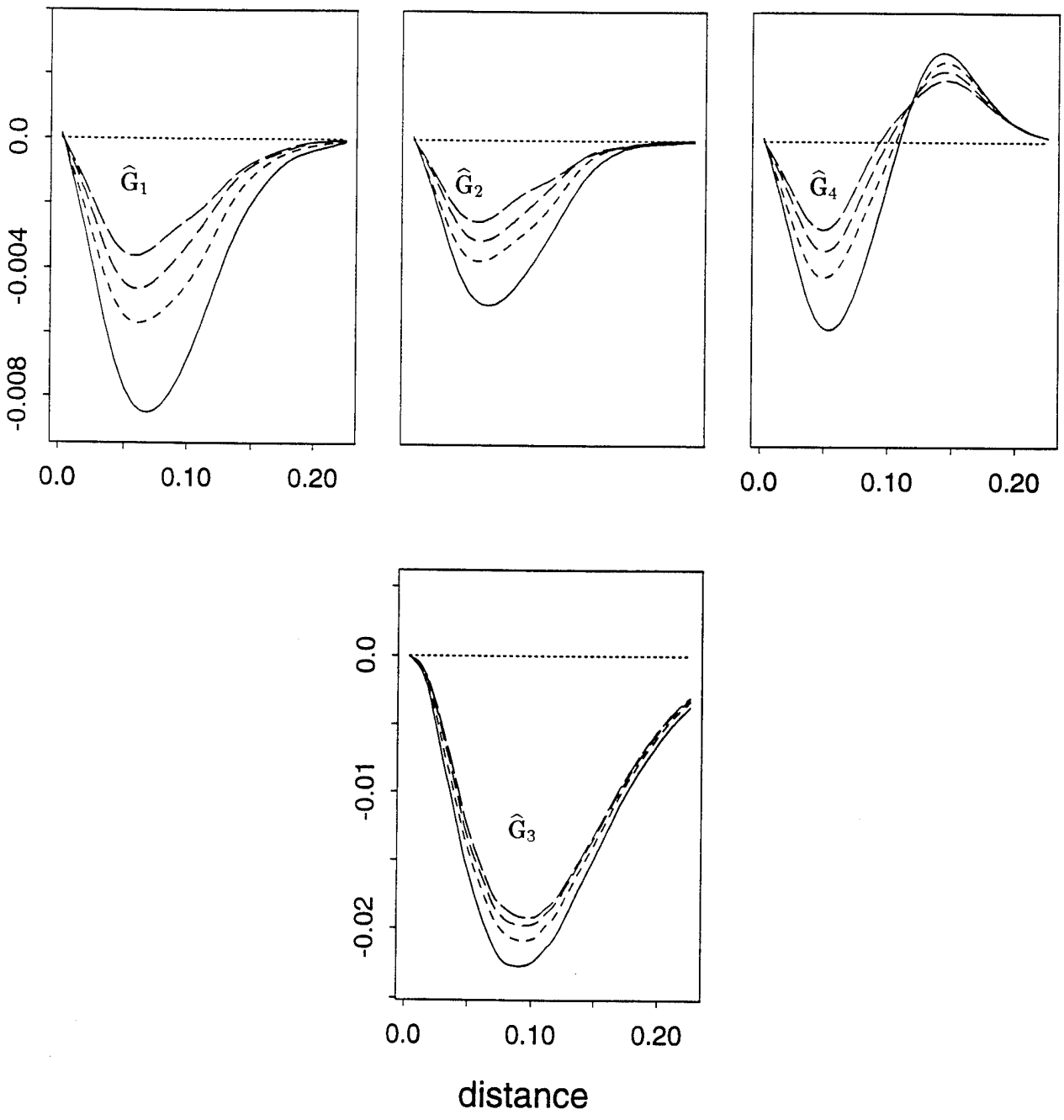
(c)



analogous at 157°

(d)

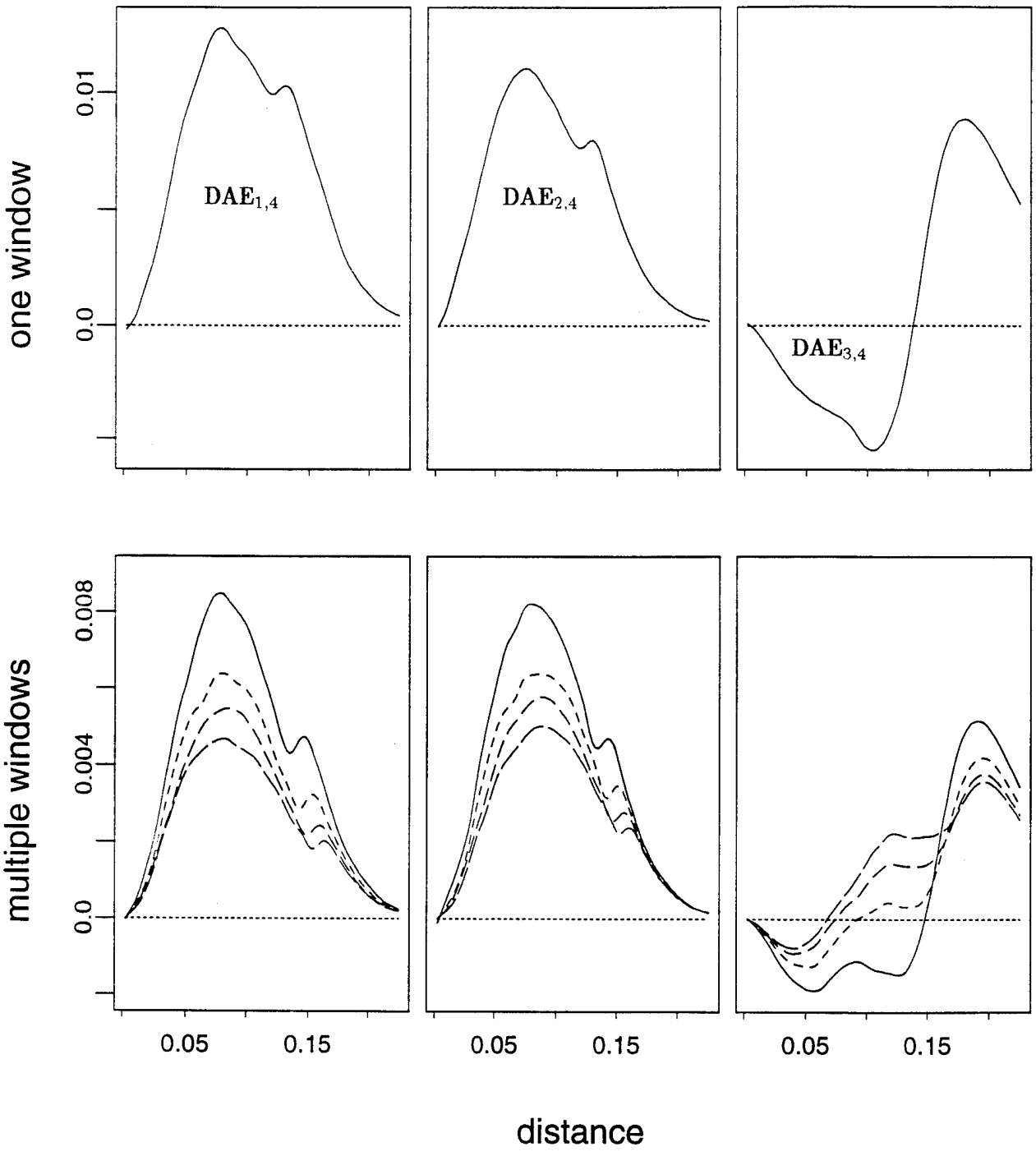
Estimated Bias

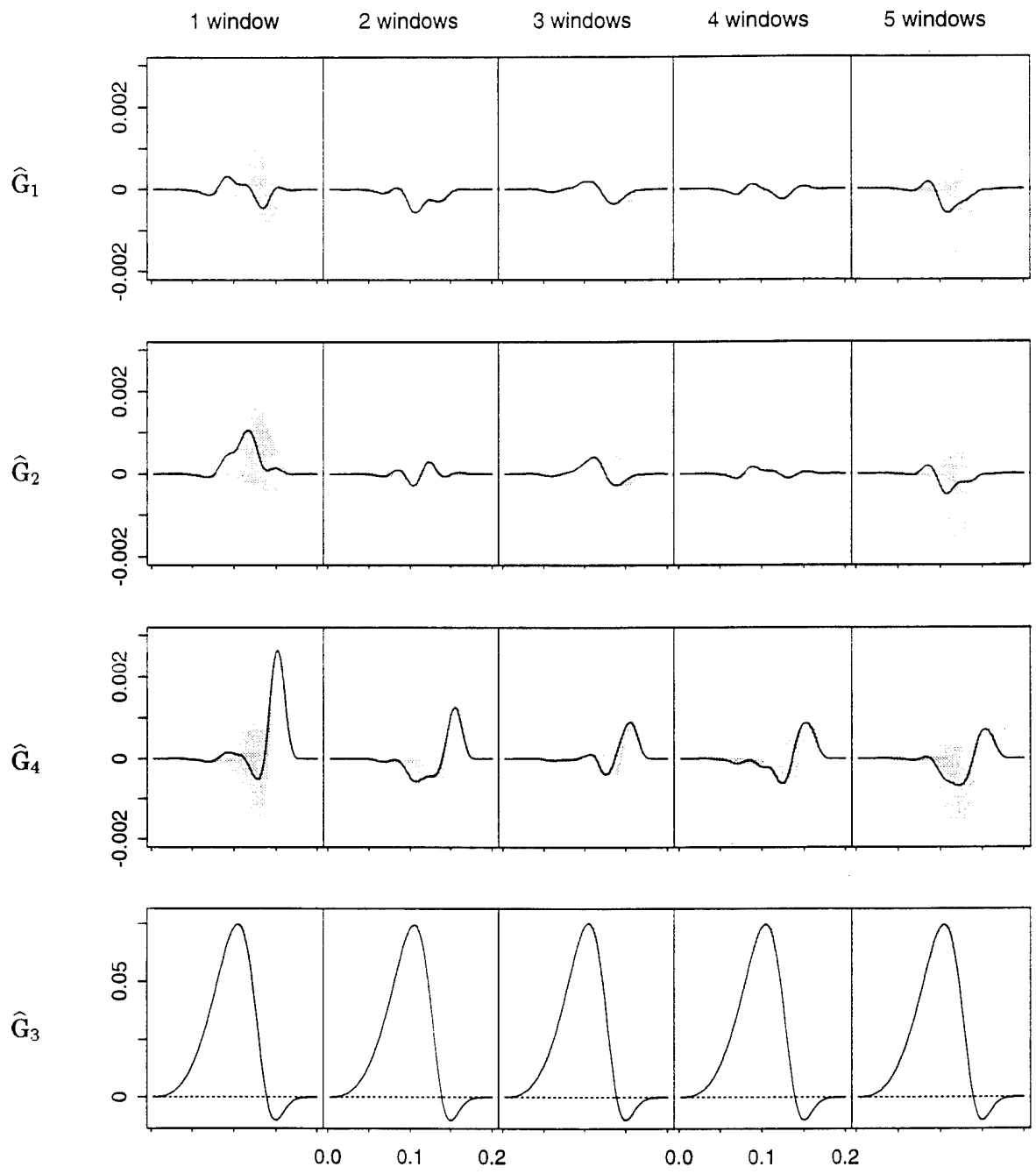


5 windows — 3 windows

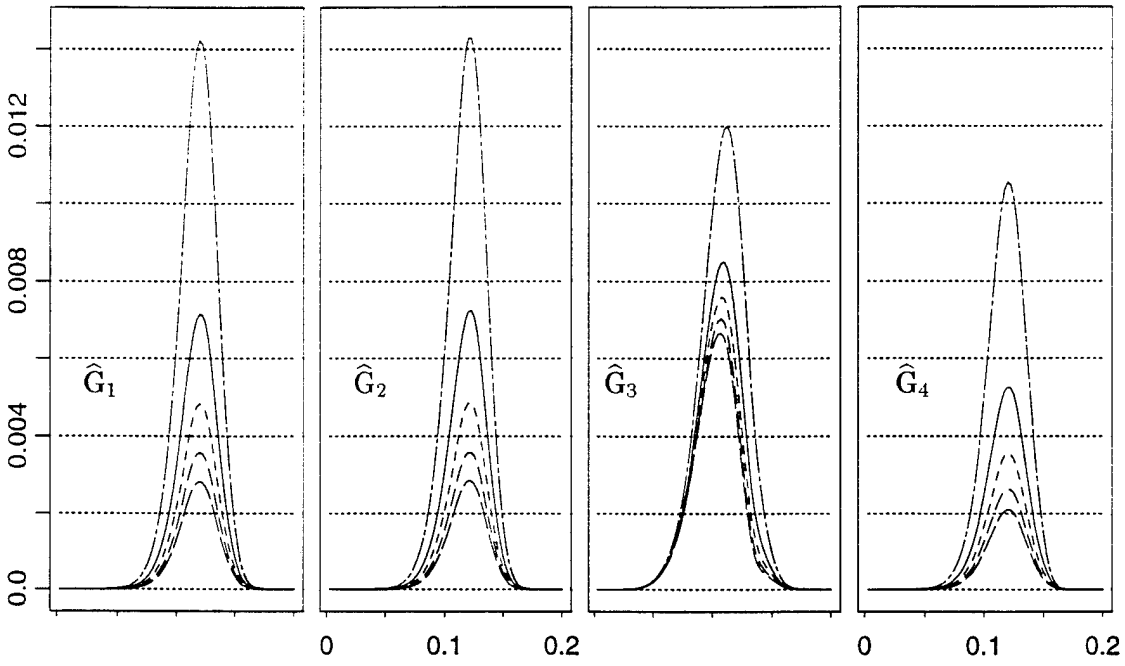
4 windows — 2 windows

Average Difference of the Absolute Errors



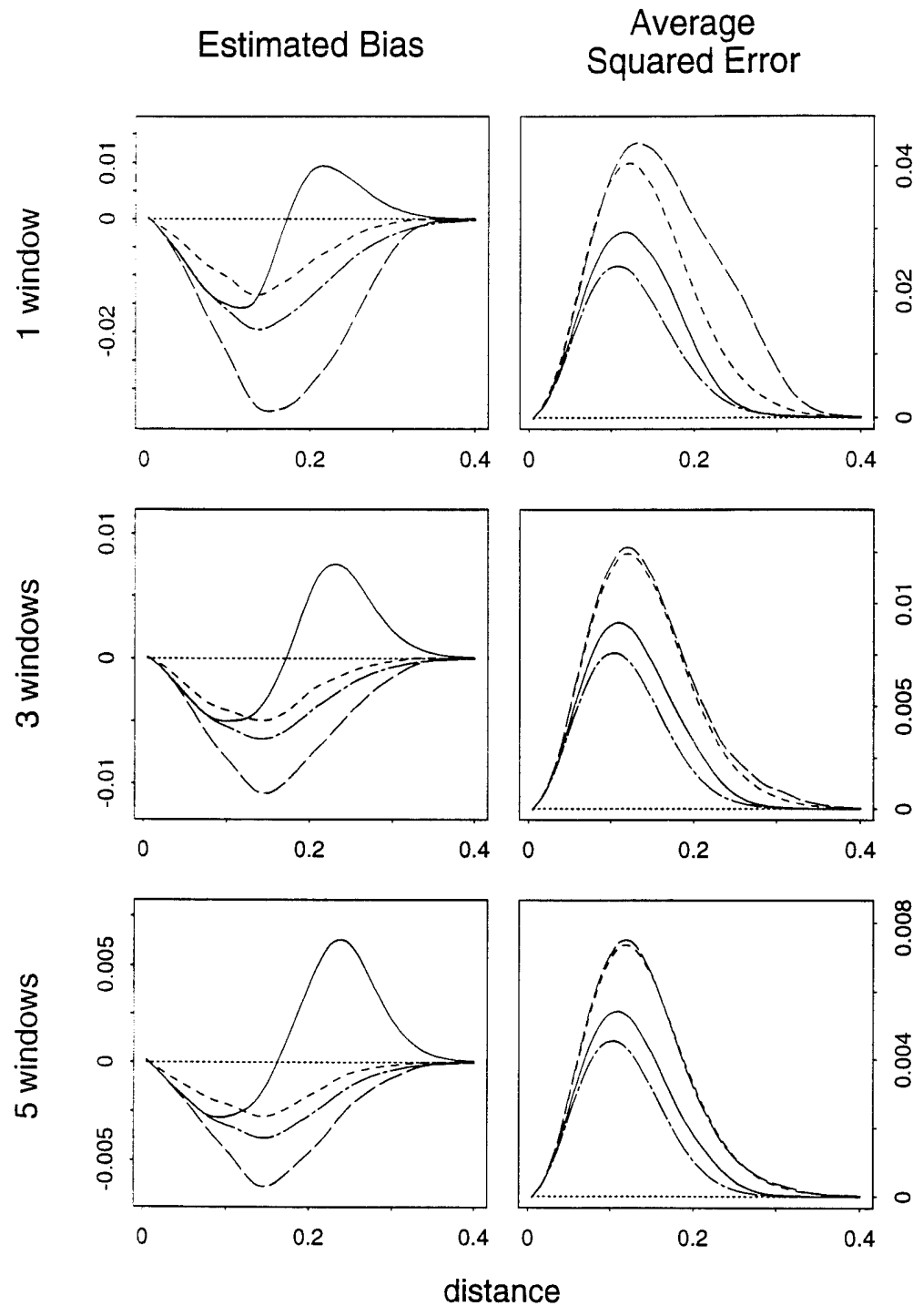

Estimated Bias
1 window
2 windows
3 windows
4 windows
5 windows
$\hat{G}_1$
$\hat{G}_2$
$\hat{G}_4$
$\hat{G}_3$
0.002
0
-0.002
0.05
0.0
0.1
0.2

Average Squared Error

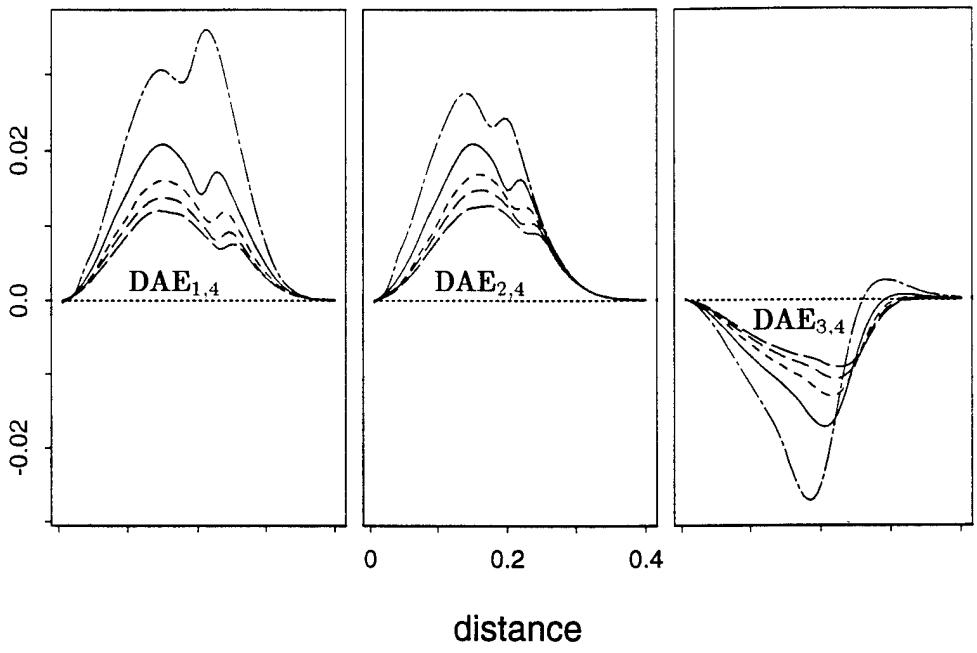


distance

5 windows
3 windows
4 windows
2 windows
1 window


Estimated Bias
Average
Squared Error
1 window
3 windows
5 windows
0.01
0
-0.02
0.04
0.02
0
0.2
0.4
-0.01
0.005
-0.005
0.008
0.004
distance
$\hat{G}_1$
$\hat{G}_2$
$\hat{G}_3$
$\hat{G}_4$

Average Difference of the Absolute Errors



5 windows
3 windows
4 windows
2 windows
1 window